Andy Zack

Exercise 4

Implementation Plan

1. Read in file
   1. Create empty list
   2. Open file
   3. Loop through lines of file
      1. If current line is not empty:
         1. Enter line as a new element in the list
2. Count lines
   1. Get length of list from step 1
3. Count words and characters
   1. Set word counter variable to zero
   2. Set character counter variable to zero
   3. Loop over elements of list from step 1
      1. Add length of each element to the character counter
      2. Add number of words (using .split()) to word counter
4. Calculate Frequency for each word
   1. Clean file by making it all lowercase and removing characters
      1. Create empty list for the cleaned lines
      2. Loop over lines in list from step 1
         1. Make current line lowercase
         2. Remove unwanted characters from current line
         3. Store current line as a new element in the cleaned list
   2. Convert list of cleaned lines to a list of cleaned words
      1. Create empty list for cleaned words
      2. Loop over lines in cleaned list
         1. Split current line into words
         2. Add current list of words to the end of the cleaned word list
   3. Create dictionary of unique words
      1. Create empty dictionary to store unique words and their frequencies
      2. Loop over words in word list
         1. If current word is not in dictionary
            1. Add current word to dictionary
            2. Set that word’s frequency to zero
   4. Add frequency counts to dictionary
      1. Loop over words in word list
         1. Increment appropriate frequency in dictionary
5. Celebrate